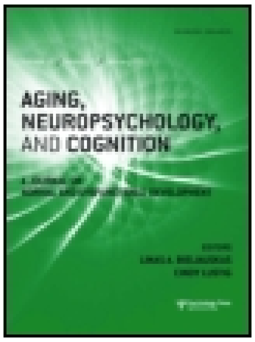


Routledge
Taylor & Francis Group

Aging, Neuropsychology, and Cognition

ISSN: 1382-5585 (Print) 1744-4128 (Online) Journal homepage: https://www.tandfonline.com/loi/nanc20

A Sensory Origin for Color-Word Stroop Effects in Aging: A Meta-Analysis

Boaz M. Ben-David & Bruce A. Schneider

Aging, Neuropsychology, and Cognition, 16: 505–534, 2009
http://www.psypress.com/anc
ISSN: 1382-5585/05 print; 1744-4128 online
DOI: 10.1080/13825580902855862

A Sensory Origin for Color-Word Stroop Effects in Aging: A Meta-Analysis

BOAZ M. BEN-DAVID AND BRUCE A. SCHNEIDER
Centre for Research on Biological Communication Systems, University of Toronto Mississauga, Ontario, Canada

ABSTRACT

An increase in Stroop effects with age is often interpreted as reflecting reductions in selective attention, or alternatively, cognitive slowing with age. In a cross-lab and a cross-sectional analysis, we linked sensory losses to Stroop declines. Specifically, we found that the latency difference, or dimensional imbalance, between reading and naming the font color of color-neutral words increased with age. A cross-sectional analysis revealed that this dimensional imbalance can both mediate the effects of age on Stroop effects, and contribute to Stroop after controlling for age effects. We conclude that age-related changes in color perception contribute to and may mediate age-related changes in Stroop.

Keywords: Cognitive aging; Stroop; Selective attention; Sensory aging; Speed of processing.

INTRODUCTION

In everyday life, one has to attend selectively to certain features in the environment while ignoring or actively suppressing others. For instance, in listening to a conversation in a restaurant, one has to focus on the words of the person sitting across the table and ignore or suppress conversations taking place at nearby tables. Age-related changes in a person's ability to selectively attend to one signal among many could potentially account for some of the difficulties that older adults experience in listening to conversations in noisy environments, or in spotting a pedestrian who is about to step off the curb in a complex visual scene.

Address correspondence to: Boaz M. Ben-David, Centre for Research on Biological Communication Systems, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, Ontario, Canada L5L 1C6. E-mail: boaz.ben.david@utoronto.ca

Are Age Differences in Stroop Interference a Reflection of Age-Related Changes in Selective Attention?

In clinical (e.g., Golden, 1978), experimental (e.g., Comalli, Wapner, & Werner, 1962), and neuropsychological (e.g., Trenerry, Crosson, DeBoe, & Leber, 1989) screening tests, age-related changes in selective attention are typically evaluated using the Stroop test. Indeed, the Stroop paradigm (Stroop, 1935) has been regarded as the golden standard of selective attention for over 70 years (see MacLeod, 1991, and Melara & Algom, 2003 for relevant reviews). In the color-word Stroop paradigm, participants are asked to name the colors of printed words, irrespective of their content. The latency advantage for naming the print color of a color-neutral word (e.g., TABLE printed in blue) over an incongruent color-word (e.g., RED printed in blue) is termed Stroop interference:

$$SI = Ci - Cn \tag{1}$$

where SI stands for Stroop interference, Ci for response times for color-naming incongruent words, and Cn for response times for color-naming color-neutral stimuli. If the participant can completely ignore or inhibit the processing of the lexical content of the colored word when asked to name the color in which it was printed, SI should equal zero. On the other hand, if she or he cannot ignore the lexical content of the word, we should observe SI > 0. Hence, investigators have assumed that an age-related increase in Ci, and consequently, an increase in the magnitude of SI, is due to a decrease in the ability to selectively inhibit lexical processing.

Indeed, when the performance of older adults (over 65 years old) is compared with that of college-aged younger adults (about 20 years old), SI was generally found to be larger for seniors (see McDowd & Shaw, 2000, for a recent review). Similarly, in life-span studies, comparing performance in different age groups ranging from college-aged to older adults, SI is found to increase with age (e.g., Comalli et al., 1962). This systematic increase of SI with age for adults is often interpreted as reflecting an age-related reduction in selective attention, consistent with Hasher and Zacks' (1988) theory of a decrease in the efficiency of inhibitory processes with aging (e.g., Hartman & Hasher, 1991; McDowd & Shaw, 2000; Troyer, Leach, & Strauss, 2006). Other investigators, however, have identified other factors that might be contributing to age differences in Stroop Interference.

Are Age-Related Differences in Stroop Interference a Reflection of Cognitive Slowing with Age?

Several authors have suggested that age-related changes in the speed of processing could serve as the source for age differences in SI. In young adults, latencies for Ci are greater than for Cn. If the speed of processing slows down with age, and latencies on all tasks are increased by the same factor with age, SI (the advantage of Cn over Ci) should increase with age. Meta-analyses of Stroop studies by Verhaeghen and De Meersman (1998) and Verhaeghen and Cerella (2002) found that Brinley plots (Brinley, 1965), in which latencies for older adults on a task are plotted as a function of latencies for younger adults on the same task, were fit by a straight line, i.e.,

$$\text{RT(old)} = a + b * \text{RT(young)},\ b > 1 \tag{2}$$

where b reflects the effects of generalized slowing of central cognitive processes. Recall that SI for younger adults is $\text{SI(young)} = \text{Ci(young)} - \text{Cn(young)}$. If Eq. (2) holds exactly for both Cn and Ci tasks, then SI for older adults is depicted as $\text{SI(old)} = b * [\text{Ci(young)} - \text{Cn(young)}] = b * \text{SI(young)}$. Since b is larger than 1, an increase of SI for older adults derives directly from Eq. (2). These considerations led Verhaeghen and colleagues (1998, 2002) to conclude that age differences in SI between younger and older adults are not the outcome of age differences in selective suppression of lexical processes, but a simple consequence of generalized slowing with age.

However, a Brinley analysis is based on a simple comparison between a group of younger and older adults on specific tasks. It does not take into account how performance actually changes with age across the lifespan. Cerella and Hale (1994; see also, Cerella, 1990), in an extensive review of aging studies, proposed a slowing model in which RTs on *a single task* increase exponentially as a function of age,

$$\text{RT} = \Psi(\text{AGE}) = \alpha * e^{\gamma * \text{AGE}} + \beta \tag{3a}$$

They further suggest that task difficulty interacted in a multiplicative fashion with cognitive slowing, to account for the variation in non-linearity between *different tasks*. Specifically, the effect of task difficulty, TD, simply multiplies the effect of cognitive slowing,

$$\text{RT} = \text{TD} * \Psi(\text{AGE}) = \text{TD} * (\alpha * e^{\gamma * \text{AGE}} + \beta) \tag{3b}$$

Note that in this model, $\Psi(\text{AGE})$ is the same for all tasks, with the function relating RT to age on one task being a simple multiplicative function of the reaction time function for any other task. This model of multiplicative cognitive slowing with age was found to explain most of the variance in life-span studies of various cognitive tasks (with adults over the age of 20). Such results are consistent with the hypothesis that the cognitive processes engaged by a task slow down with age at the same rate, and that rate differences across tasks are due to the multiplicative effect of task difficulty. In the cross-sectional analysis part of this paper, we test the Cerella and Hale model with data from a single life-span study.

Are Age-Related Differences in Stroop Interference a Reflection of Age-Related Sensory Decline?

Melara and Algom (2003) proposed that when access to the lexicon is faster than access to the font color of the word, Stroop interference ensues. For example, when asked to name the font of an incongruent color word (RED printed in blue), participants have to inhibit the response to the name of an incongruent word ('RED') until access is gained to the color in which it was printed and the correct color response is activated ('blue'). A failure to do so presumably produces Stroop effects. The greater the difference in the speed of processing in the two dimensions (accessing and naming the font color versus accessing the lexicon and reading the word), the greater is the resultant Stroop interference. Melara and Algom referred to this difference in accessibility as dimensional imbalance and measured it by computing the latency difference between naming the font color of color-neutral stimuli and reading color-neutral stimuli (relative baseline accessibility, their Eq. 5, p. 432):

$$\mathrm{DI} = \mathrm{Cn} - \mathrm{Rn} \tag{4}$$

where DI stands for dimensional imbalance, and Cn and Rn stand for latencies of color-naming and reading color-neutral stimuli, respectively. As DI increases, the time difference between access to the lexical content of an incongruent word and access to its print color will increase, leading to a larger SI. Indeed, in an analysis of 34 experiments, Melara and Algom found a significant positive regression (r = .78) between dimensional imbalance and Stroop effects, larger imbalance scores were accompanied by larger Stroop effects (see their Figure 3b, p. 429). Hence, if age differentially affected access to the name of the word versus the print color, we would expect age-related changes in DI, and as a result, age-related changes in SI. Changes in DI would result in changes in SI, independent of any contribution of age-related changes in the ability to selectively inhibit lexical processing (the selective attention explanation for SI), or of generalized age-related slowing in cognitive processes (the generalized slowing explanation for SI).

A recent study, Bugg, DeLosh, Davalos, and Davis (2007), presents evidence that generalized slowing may not be the sole explanation for age differences in Stroop performance. In this study, 938 participants, aged 20–89, completed an abbreviated Golden color-word Stroop task (Golden, 1978). The authors found that 74% of the variance in Ci remained unaccounted for after controlling for speed of processing as measured by Cn, with age accounting for significant additional variance. With a subset of 281 participants, 78 and 66% of the age-related variance in Ci was unaccounted for after controlling for speed of processing as measured by simple and choice reaction tasks, respectively. The authors conclude that: 'further study is needed, however, to better understand the unique effects of age on Stroop interference, beyond the influence of general slowing' (p. 166).

A possible source of additional effects of age on Ci beyond speed of processing is suggested by the results of a study by Salthouse and Meinz (1995). They conducted a cross-sectional study of 14 cognitive tasks (including Ci, Cn and Rn) using 5 age groups from 20 to 89 years old. Then, they computed the proportion of age-related variance shared between Ci and each of the 13 other tasks. They found that the age-related shared variance between Ci and 12 of these tasks ranged between 72 and 88% (the highest shared variance was with Cn). However, the amount of age-related shared variance between Ci and Rn was only 38% (similarly, the smallest amount of shared age-related variance with Cn was with Rn). A possible reason for this is that age had a much smaller effect on reading than on any other examined task (see their Figure 3, p. 303). The authors concluded that 'smaller age relation for reading speed measures than for other naming measures . . . may be related to the extensive experience most people have had reading words' (p. 305). This discrepancy or 'imbalance' between a substantial age-effect on color-naming speed with a much milder effect on reading speed can be seen as another possible source for age differences in Stroop effects, above and beyond a model of generalized cognitive slowing coupled with differences in task difficulty.

A glance at the pertinent literature suggests more reasons why DI may change with age. Color vision deteriorates rapidly after the age of 60: the change emanates mostly from a yellowing of the lens and from a general reduction in the number of photo receptors, with the largest reduction occurring in the short wavelength (blue) system (e.g., Werner & Steele, 1988; Nguyen-Tri, Overbury, & Faubert, 2003). Anstey, Dain, Andrews, and Drobny (2002), in a study of older adults aged 60–87 years, found that latencies for Ci were correlated (r = .52) with a measure of color-vision (Farnsworth–Munsell Panel Test, D15, Farnsworth, 1943). In a structural equation model, color-vision was found to explain more of the age-related variance in Ci than any other variable. The authors conclude that 'a significant proportion of the observed age differences in Stroop may be attributed to individual differences in visual abilities' (p. 262).

Word-reading latencies are also known to change with age (e.g., Rodriguez-Aranda, 2003). However, some studies show that these age changes may be very limited. For example, Akutsu, Legge, Ross, and Schuebel (1991) showed that healthy older adults (with no ocular disease) read as fast as younger adults when letter sizes were optimal (0.3–1.0 visual degrees, see their Figure 4, p. 328). As seen earlier, Salthouse and Meinz (1995) also found that reading was slowed down with age much less than color-naming (of either neutral, Cn, or incongruent stimuli, Ci). If age-related visual declines take a larger toll on color-naming than on reading, DI will increase with age, and lead to greater Stroop effects. In other words, we suggest that a sensory source – deterioration in color vision with age – may contribute to age-related changes in Stroop effects.

The goal of this study is to explore the possibility of a sensory origin for aging differences in Stroop effects, and to compare it to the cognitive slowing account and the traditional selective attention deficit account for age-related Stroop differences. In the first step, we compared paired groups of younger and older adults taken from 13 different studies, in a *cross-lab analysis* of age-related changes in Stroop effects. In the second step, we conducted a detailed *cross-sectional analysis* of a single study (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006) in which Stroop performance was measured in 12 different age groups ranging from 25 to 80 years. In the third step, we fit different models to the data from the latter study.

METHOD

Sample of Studies

Studies were collected (in October 2008) by consulting the Scopus electronic database,[1] and by reviewing the references in the retrieved articles. We selected only studies that included: (a) data for both a group of younger (under 25 years old) and older (over 65 years old) adults and (b) measures of Rn, Cn, and Ci. Rn measures could be either reading aloud color-neutral words printed in color (e.g., the word TABLE in red font) or reading aloud color-words printed in black (e.g., the word RED printed in black on a white background). Cn measures could be naming aloud the font color of neutral words (e.g., the word TABLE in red), naming aloud the font color of strings of X's printed in color, or naming aloud the color of color patches. The Ci measure in all studies was naming aloud the font color of incongruent color-words (e.g., the word RED printed in blue). Because the type of response (oral or button press) differentially affects Rn, Cn, and Ci (for a review, see Melara & Algom, 2003; also see Logan & Zbrodoff, 1998), we limited the analysis to only those studies in which there was an oral response in all three tasks. In total,

[1]We have complimented this search by cross-searching Pub Med and PsychInfo electronic databases as well.

TABLE 1. The Characteristics of each Age Group and the Mean Latencies (in seconds) for Reading a Neutral Word, Rn, Color-Naming a Neutral Stimulus, Cn, and Color-Naming an Incongruent Word, Ci, as used in the Cross-Laboratory Analysis

	Younger Adults					Older Adults				
	Age	*N*	Rn	Cn	Ci	Age	*N*	Rn	Cn	Ci
Bialystok et al. (2008)	20	48	0.527	0.598	0.741	68	48	0.566	0.642	0.862
Bugg et al. (2007)	26	443	0.404	0.455	0.949	71	212	0.444	0.554	2.478
Comalli et al. (1962)	18	18	0.405	0.561	1.030	72.5	15	0.451	0.689	1.651
Dulaney and Rogers (1994)	22	40	0.387	0.519	0.789	70	40	0.469	0.674	1.212
Hartman and Hasher (1991)	20	44	0.399	0.560	0.993	66.5	24	0.404	0.609	1.275
Houx et al., (1993)	20.5	22	0.401	0.511	0.773	70	25	0.413	0.564	0.995
Kieley and Hartley (1997)	22	16	0.511	0.557	0.736	75	16	0.520	0.593	0.866
Klein et al. (1997)	30	121	0.420	0.540	0.820	74	71	0.500	0.680	1.240
Rodriguez-Aranda and Sundet (2006)	26	25	0.434	0.641	1.004	71	50	0.515	0.781	1.554
Spieler, Balota, and Faust (1996)	20.5	27	0.519	0.671	0.759	70.5	25	0.635	0.894	1.069
Uttl and Graf (1997)	25	63	0.385	0.494	0.717	70	38	0.432	0.577	0.933
Van der Elst et al. (2006)	25	159	0.396	0.525	0.795	70	155	0.460	0.608	1.108
West and Baylis (1998)	22.5	40	0.391	0.532	0.758	69.5	40	0.422	0.666	1.288

we collected data from 13 published studies. A list of these studies, along with some of their characteristics, can be found in Table 1.

Data Pooling for the Cross-Lab Analysis

We chose from these 13 studies groups of younger adults, under the age of 30 years old and over 18, and groups of older adults, under the age of 75 and over 65 (young–old), subject to the above constraints. These groups include 1825 participants: 1066 younger adults (average age 25 years) and 759 older adults (average age 70.5 years), after excluding sub-groups of participants with reported serious health problems (e.g., the 'biological life events – BLE' group in Houx, Jolles, & Vreeling, 1993). Several discrepancies between the data presented in Table 1 in our study and the data used in the meta-analysis by Verhaeghen and De Meersman (1998, Table 1, p. 121) arise from the above-listed constraints (for example, Verhaeghen & De Meersman, included data presented in the BLE group in Houx et al., 1993). In two of the selected studies, the data were read off the graph (Comalli et al., 1962, Figure 1; and Klein, Ponds, Houx, & Jolles, 1997, Figure 1a). In three other studies (West & Baylis, 1998, Figure 1; Bugg et al., 2007, Figure 1; and Bialystok, Craik, & Luk, 2008, Figure 2) the exact data were received in personal correspondence with the authors. In each experiment, the data were averaged to present the mean response time per a single item. Finally, in

FIGURE 1. Age differences, RT(old) – RT(young), in response latencies for reading neutral words as a function of response latencies for color-naming neutral stimuli in 13 color-word Stroop studies. The solid line is the least-squares regression line.

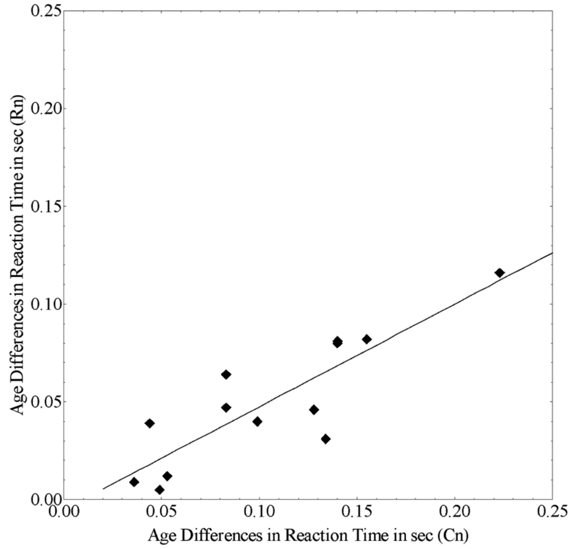
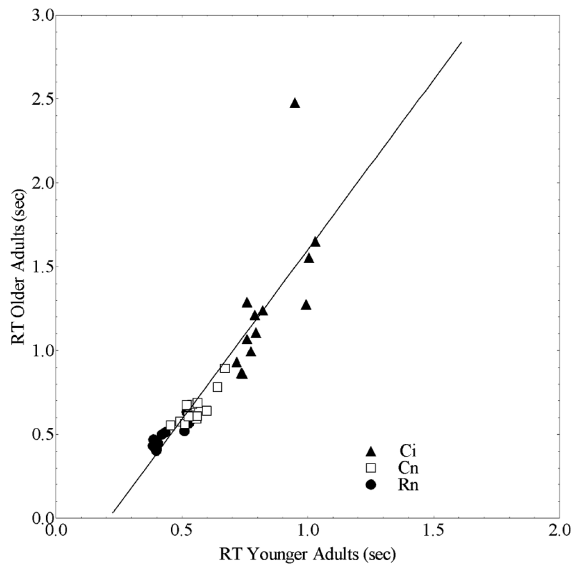


FIGURE 2. Brinley plot (latencies for older adults as a function of latencies for younger adults) of color-naming incongruent words (Ci, triangles), color-naming neutral stimuli (Cn, unfilled squares) and reading neutral words (Rn, circles) tasks in 13 color-word Stroop studies, along with the least squares regression line (solid line).



three studies (Dulaney & Rogers, 1994; Van der Elst et al., 2006; Bialystok et al., 2008) the data presented in Table 1 is a weighted average of different experimental groups that come from the same age cohort.

Cross-Sectional Analysis of a Single Study

An analysis across the life-span of the Stroop task was conducted on the data taken from the Van der Elst et al. (2006) study. This extensive study presents the average performance of 1788 participants in 12 age categories (25–81 years). Each group is further divided into six sub-groups according to gender and three levels of obtained education (low, average and high). For this analysis, we computed a weighted average across the specific subgroups (taken from their Table 4, pp. 69–71).

RESULTS

A Cross-Laboratory Analysis of 13 Studies

Age Differences

In this part of the analysis, we compared the mean performance of groups of older adults and groups of younger adults in 13 different studies. All of these studies reported a significant age difference in Stroop effects. It is not surprising to find that both SI and latencies for Ci increased with age in all of these 13 studies [$t(12) = 3.35$, $p = .006$ and $t(12) = 4.31$, $p = .001$, respectively; Wilcoxon $Z = 2.803$, $p = .005$, for both]. However, it is a novelty to find that DI scores were also higher for older than for younger participants in all studies [$t(12) = 6.43$, $p = .00003$; Wilcoxon $Z = 2.803$, $p = .005$]. To evaluate whether age slows down color-naming more than it does reading, we plotted, in Figure 1, age differences in Rn as a function of age difference in Cn, across the 13 studies. The two were highly correlated [$r^2 = .745$, $F(1, 11) = 32.07$, $MSE = 0.0003063$, $p = .0002$] with a slope of 0.53 and an intercept non-significantly different than zero [$t(11) = –0.648$, $p > .6$], indicating that age differences in reading speed were about half of those for color-naming speed, hence, the increase in DI for older adults. These results may indicate that older adults, when naming the font color of incongruent color-words, would have to inhibit the response to reading for a much longer period than younger adults, thus generating larger Stroop effects, and slower Ci latencies.

Brinley Plots

Figure 2 plots the reaction times of older adults as a function of the reaction time for younger adults for Ci (triangles), Cn (squares), and Rn (circles) conditions (a Brinley plot). An examination of this plot shows that Rn response times were faster than Cn, which in turn were faster than Ci. To investigate whether the slope of the line in this Brinley plot differed in three tasks (Cn, Ci, and Rn), we fit the following regression equation to the data:

$$\begin{cases} \mathrm{Cn(old)} = a_\mathrm{n} + b_\mathrm{n} * \mathrm{Cn(young)} \\ \mathrm{Ci(old)} = a_\mathrm{i} + b_\mathrm{i} * \mathrm{Ci(young)} \\ \mathrm{Rn(old)} = a_\mathrm{R} + b_\mathrm{R} * \mathrm{Rn(young)} \end{cases} \quad (5)$$

We compared this six-parameter model to a two-parameter model in which $a_\mathrm{n} = a_\mathrm{i} = a_\mathrm{R}$, and $b_\mathrm{n} = b_\mathrm{i} = b_\mathrm{R}$, thus the same slope and intercept are used for each of the three conditions. Increasing the number of parameters from 2 to 6 did not result in a significant reduction in variance [$F(4, 33) = 1.00$, $MSE = 0.0362$, $p > .4$]. Therefore, a single linear regression line, with a single slope and a single intercept [$a = 2.026$, and $b = -0.4239$; $r^2 = .807$, $F(1, 37) = 154.56$, $MSE = 0.036195$, $p < .00001$] is sufficient to describe the data for Cn, Ci and for Rn. To estimate the percentage of variance in the population accounted for by a model in which the response times of older adults on each of the three tasks were linearly related to the response times of younger adults ($a_\mathrm{n} = a_\mathrm{i} = a_\mathrm{R}$, and $b_\mathrm{n} = b_\mathrm{i} = b_\mathrm{R}$ in Eq. 5), we computed the adjusted r^2 statistic (adj-r^2). This statistic takes into account the effects of the sample size and of the number of predictor variables when estimating the proportion of the population variance accounted for in the model. The adj-r^2 for this simple linear model was .802 indicating that this model accounts for a large proportion of the variance in the linear function relating the response times of older adults to those of younger adults. Note that age differences in selective attention would predict a different slope for Ci than for Cn or Rn, because the Ci task requires inhibition (of the lexical dimension), whereas none is required in Cn or in Rn. Hence, there is no evidence from these 13 studies to support the notion that age-related differences in selective attention are responsible for age differences in SI, a conclusion also reached by Verhaeghen and De Meersman (1998) and Verhaeghen and Cerella (2002). Note that Figure 2 implies that DI will increase with age at the same rate as any other task. This follows from the fact that if $a_\mathrm{n} = a_\mathrm{R}$ and $b_\mathrm{n} = b_\mathrm{R}$, then $\mathrm{DI(old)} = b * [\mathrm{Cn(young)} - \mathrm{Rn(young)}] = b * \mathrm{DI(young)}$.

Generalized Slowing or Dimensional Imbalance?

The results to this point are compatible with the hypothesis that age differences in Stroop effects are a function of generalized slowing. However, because we also found age differences in DI, the results are also compatible with the hypothesis that age differences in Stroop effects are due to age differences in DI. Support for the direct contribution of DI to Stroop effects comes from studies by Algom, Dekel, and Pansky (1996), and Sabri, Melara, and Algom (2001). Algom and colleagues have shown that direct experimental manipulations of DI lead to changes in Stroop effects, with Stroop effects increasing with DI. Hence, it is reasonable to hypothesize that the effects of age on SI may be mediated, in part, by age differences in DI. On the other hand, Figure 2 suggests that age changes in DI may be a simple consequence of generalized slowing. If so, then generalized slowing could account for all of the variance in SI. To support the argument that age differences in DI contribute to age differences in SI, we examined, in the next section, whether DI could account for some of the residual variance in Ci, once the effects of age had been removed. In order to investigate this possibility, we turn to the most extensive study of age-related changes in Stroop effects across the lifespan that satisfied the criteria listed in the Method section, namely the Van der Elst et al. (2006) study.

A Cross-Sectional Analysis of a Single Study

In the previous cross-lab analysis we compared studies that used different apparati, different stimuli for the color-neutral conditions (e.g., color patches, strings of X's printed in color or color-neutral words printed in black), different stimulus presentation modes (computer or cards), and different viewing conditions (font, font sizes, font and background colors). Melara and Algom (2003) argued that these changes might affect both DI and Stroop effects. One could also argue that different versions of the same task are likely to differ in task difficulty which, according to the Cerella and Halle (1994) model (Eq. 3b), would lead to variations in reaction times. In any event, performance variability is likely to be fairly large, because the participants were tested under different conditions in the various laboratories, and may have been selected from different populations (especially, the groups of older adults). Hence, it is possible to assume that the lack of any differences in slopes among the three tasks in a Brinley analysis, across different studies, may reflect a lack of statistical power that is inherent to this type of analysis. In this section, we avoided these additional sources of variance, by performing a fine-grained analysis of the change in performance over the life-span in a single extensive study, Van der Elst et al. (2006). It is noteworthy that all of the stimuli in this study were presented on cards, and participants were asked to read aloud the word, or name aloud the color of each stimulus while timed. The color-neutral stimuli for Cn were color patches that do not carry any lexical meaning,[2] and for Rn, they were

[2]This was a possible confound in some of the studies examined earlier which used color-neutral words that could have possibly interfered with the process of color-naming (e.g., see Klein, 1964; Monsell, Taylor, & Murphy, 2001).

TABLE 2. Mean Latencies (in seconds) for Reading a Neutral Word, Rn, Color-Naming a Color-Patch, Cn, and Color-Naming an Incongruent Word, Ci, from the Van der Elst et al. (2006) Study, Averaged Across Subgroups in Each Age Cohort, as used in the Cross-Sectional Analysis

Age	*N*	Rn	Cn	Ci
25	159	0.396	0.525	0.795
30	154	0.410	0.524	0.810
35	157	0.399	0.515	0.804
40	155	0.412	0.528	0.833
45	162	0.415	0.530	0.863
50	160	0.429	0.550	0.890
55	157	0.434	0.571	0.976
60	157	0.453	0.590	0.997
65	153	0.460	0.610	1.063
70	155	0.460	0.608	1.108
75	157	0.492	0.658	1.242
80	59	0.499	0.685	1.409

color-words printed in black on a white background. The data used for this analysis are presented in Table 2.

Age Effects

Figure 3 plots mean reaction times as a function of age groups for the three tasks, Rn (triangles), Cn (squares), and Ci (circles). This figure indicates that reaction times appear to be non-linearly related to age on all tasks. To test this, we first examined whether reaction times for each of the three tasks (Rn, Cn and Ci) were non-linearly related to age, by determining whether a quadratic model,

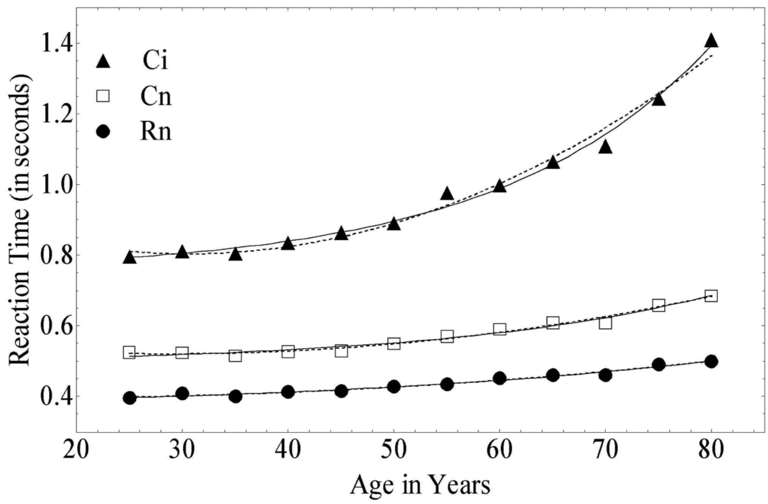
$$\mathrm{RT} = d_0 + d_1 * \mathrm{AGE} + d_2 * \mathrm{AGE}^2 \tag{6a}$$

provided a significantly better fit to the data than a linear model,

$$\mathrm{RT} = d_0 + d_1 * \mathrm{AGE} \tag{6b}$$

For all three tasks, the quadratic model (Eq. 6a) significantly improved the percentage of variance accounted for by age [for Rn, $F(1, 9) = 12.64$, $MSE = 0.0000397$, $p < .01$; for Cn, $F(1, 9) = 41.63$, $MSE = 0.0000764$, $p < .0005$; for Ci, $F(1, 9) = 57.31$, $MSE = 0.0007643$, $p < .0001$]. Hence, Rn, Cn, and Ci were non-linearly related to age. For Rn, adding the quadratic term increased

FIGURE 3. Latencies of color-naming incongruent words (Ci, triangles), color-naming neutral stimuli (Cn, unfilled squares) and reading neutral words (Rn, circles) as a function of age, taken from Van der Elst et al. (2006). The dashed curves represent the second order regression fit for each task (Eq. 6a), and the smooth curves represent the exponential regression fit (Eq. 3a).



the effect size (adj-r^2) from 93 to 97%; for Cn, from 88 to 98%; and for Ci from 87 to 98%. Hence, adding the quadratic term to the linear regression of reading on age increased the estimate of the population variance accounted for by this variable by only 4%, whereas adding a quadratic term for color-naming neutral and color-naming incongruent words increased the adj-r^2 by 10 and 11%, respectively.We also tested whether any significant improvement in fit could be obtained by adding a cubic age term,

$$RT = d_0 + d_1\text{AGE} + d_2\text{AGE}^2 + d_3\text{AGE}^3 \qquad (6c)$$

We found that adding a cubic age term did not provide any further significant improvement on any of the three tasks [for Rn, $F(1, 8) = 0.043$, $MSE = 0.0000444$, $p > .5$; for Cn, $F(1, 8) = 0.39$, $MSE = 0.0000819$, $p > .5$; for Ci, $F(1, 8) = 5.1$, $MSE = 0.0005249$, $p > .05$]. Hence, a quadratic relationship (Eq. 6a) provides a good description of the effect of age on Rn [adj-$r^2 = .967$, $F(2, 9) = 160.36$, $MSE = 0.0000397$, $p < .00001$], Cn [adj-$r^2 = .976$, $F(2, 9) = 225.84$, $MSE = 0.0000764$, $p < .00001$] and on Ci [adj-$r^2 = .980$, $F(2, 9) = 266.07$, $MSE = 0.0007643$, $p < .00001$].

Following Cerella and Hale (1994) we also examined the exponential relations between age and Rn, Cn, and Ci as depicted in Eq. (3a), by minimizing the SSE in each relation. The exponential model accounted for 97.2, 97.5 and

99.2% of the variance in these variables, respectively. The equivalent quadratic fits (listed above) accounted for 97.3, 98 and 98.3% of the variance, respectively. Both fits are shown in Figure 3. Hence, either way of fitting the non-linear relationship accounts for an equivalent amount of variance. In sum, all three tasks increase with age in a non-linear fashion that can be described by either a quadratic equation or an exponential function.

Different Slowing Rates of Cn and Rn Result in Age-related Increase in DI

The best fitting quadratic functions to Ci, Cn, and Rn are shown in Figure 3. A close examination of these functions suggests that the rate of growth in reaction time with age may differ across the three tasks. To test this, we compared a model in which we fit Eq. (6a) with the same rate of increase for all three tasks (the same d_1 and d_2 for all tasks), but with different intercepts (d_0) to a model in which individual functions were fit to the separate tasks (different rates of increase and different intercepts). An F-test showed that the full model (individual fits of d_0, d_1, and d_2) in the three tasks provided a significantly better fit than the reduced model [$F(4, 27) = 138.25$, $MSE = 0.0002935$, $p < .00001$]. We repeated this procedure, comparing the reduced model to the full model, with the pairs Rn:Cn, Rn:Ci, and Cn:Ci. In all of these pairwise comparisons, the full model provided a better fit to the data than the reduced model: [$F(2, 18) = 40.02$, $MSE = 0.0000348$, $p < .00001$, for Rn and Cn; $F(2, 18) = 287.1$, $MSE = 0.0002412$, $p < .00001$ for Rn and Ci; $F(2, 18) = 202.6$, $MSE = 0.0002522$, $p < .00001$, for Cn and Ci]. Hence, all three tasks differed significantly from each other with respect to growth rate, with the slowest growth for Rn and the fastest for Ci. The effect size (adj-r^2) for the full model in which the rate of growth in reaction time with age differs across the three tasks is 0.999. Figure 4 plots the predicted reaction times based on the full and reduced models for each task separately. As Figure 4 shows, the reduced model predicts a growth function for Rn that is much steeper than that of the actual data points (in fact, the mean reaction time provides a better fit to the data points than does the reduced model). It also predicts a growth function for Cn that is steeper than that of the data points. This growth function can account for only 41% of the variance in Cn scores. Finally, the reduced model predicts a growth function for Ci that is shallower than the obtained data, and accounts for only 73% of the variance in the Ci reaction times. Note, however, as indicated by the adj-r^2, the full model provides an excellent fit to the data points.

The data shows that latencies for naming the color of neutral words and for reading these words slowed down with age at different rates. Hence, as Figure 5a shows, DI, the advantage of Rn over Cn (Eq. 4), appears to increase as a function of age. Because we have argued that DI may mediate between age and Ci, we also checked to see if DI was non-linearly related to age. We found that adding the quadratic term significantly improved the fit

FIGURE 4. The predictions of a model which allowed for different growth rates as a function of age (solid line, Eq. 6a, allowing for different d_1, d_2 and d_3) for reading neutral words (Rn, panel A), for color-naming neutral stimuli (Cn, panel B), and for color-naming incongruent words (Ci, panel C) versus a model in which the growth rates were constrained to be equivalent on all three tasks (dashed line Eq. 6a, with the same d_1, d_2 for all three tasks). Data from Van der Elst et al. (2006).

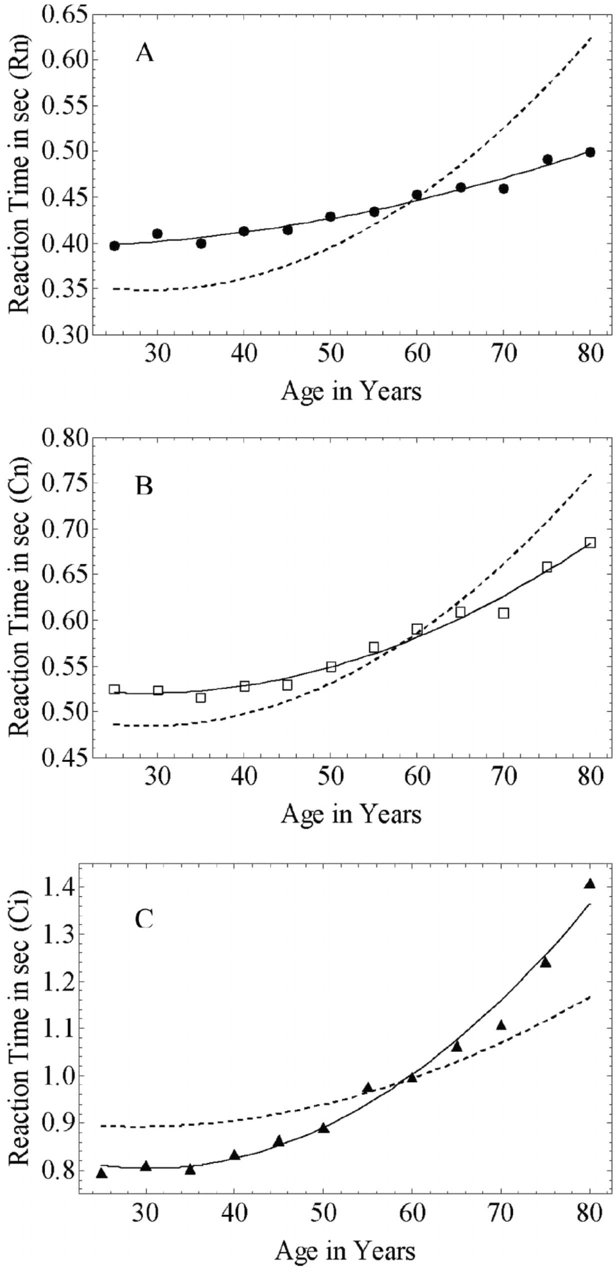
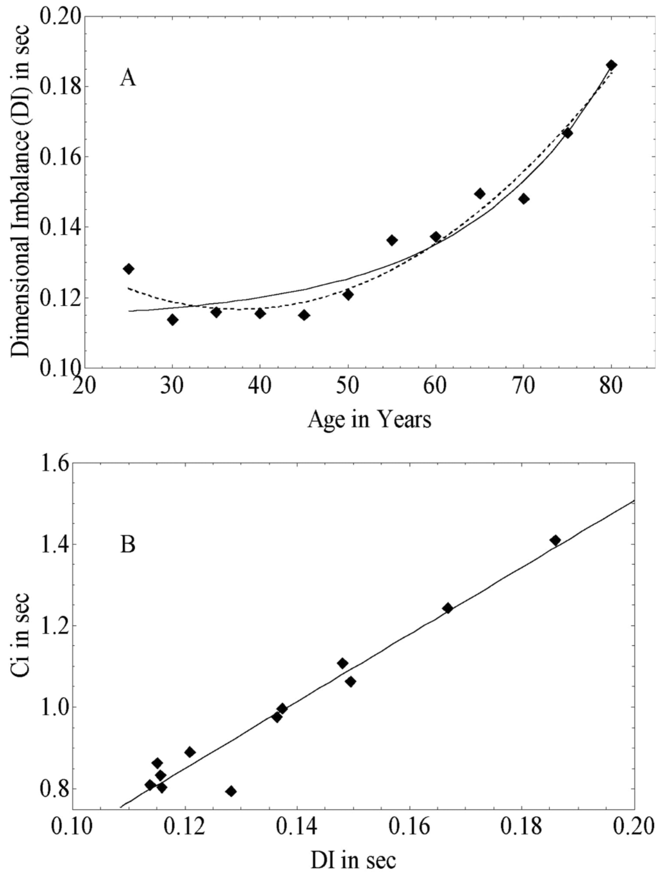


FIGURE 5. Data from Van der Elst et al. (2006). (A) Dimensional imbalance, DI, (latency advantage for color-naming over reading neutral stimuli) as a function of age. The dashed curve represents a quadratic relation (Eq. 6a) and the smooth curve represents an exponential fit to the data (the first line of Eq. 8); (B) color-naming incongruent words (Ci) as a function of dimensional imbalance (DI). The straight line represents a linear fit to the data (the second line of Eq. 8).



between DI and age [$F(1, 9) = 43.06$, $MSE = 0.0000269$, $p < .0005$], but that adding a cubic term did not significantly improve the fit [$F(1, 8) = 0.6488$, $MSE = 0.0000280$, $p > .05$]. The effect size (adj-r^2) for the quadratic relation between age and DI is 0.949 [$F(2, 9) = 103.75$, $MSE = 0.0000269$, $p < .00001$]. An exponential fit of the relation between DI and age, as depicted in Eq. (3a), was found to explain 93.6% of the variance. Figure 5A presents both the quadratic (continuous line) and the exponential (dashed line) functions of DI as a function of age.

Figure 5B suggests that Ci is linearly related to DI. This was confirmed by a statistical analysis that showed that adding a quadratic term to the linear relationship did not significantly improve the fit [$F(1, 9) = 1.06$, $MSE = 0.0021557$, $p > .1$]. Hence, Ci is linearly related to DI [adj-$r^2 = .942$, $F(1, 10) = 180.71$, $MSE = 0.0021688$, $p < .00001$, slope = 8.2], with an intercept that is not significantly different than zero [$t(10) = 1.6$, $p > .1$]. In sum, the dimensional imbalance between the speed of color naming and the speed of reading increaser with age; and Stroop performance (in Ci) is linearly related to this increase.

Model Testing

Does DI have a Residual Effect on Ci after Controlling for Age?

So far, we have shown that both age and DI are strongly related to Ci. Therefore, in this section we examined whether adding DI (or Cn, or Rn) to age in a prediction equation could significantly increase the amount of accounted for variance in Ci, above that which could be attributed to age. In other words, does adding DI (or Cn or Rn) have any residual effect on Ci after controlling for age. In Eq. (7a), adding linear and quadratic DI terms to Eq. (6a) did significantly increase the accounted for variance in Ci [$F(2, 7) = 25.6$, $MSE = 0.0001182$, $p < .001$].

$$\text{Ci} = d_0 + d_1 * \text{AGE} + d_2 * \text{AGE}^2 + f_1 * \text{DI} + f_2 * \text{DI}^2 \tag{7a}$$

Adding Rn terms instead of DI to Eq. (7a) did not account for significantly more variance than age alone [$F(2, 7) = 1.53$, $MSE = 0.0006834$, $p > .05$]. Adding Cn to Age did [$F(2, 7) = 7.79$, $MSE = 0.0003048$, $p < .05$], but not to the same extent as DI. Hence, adding DI to age as a predictor of Ci increased the effect size (adj-r^2) from 98.0 to 99.7% (an increase of 1.7%), whereas adding Cn to age increased adj-r^2 by 1.2%. Therefore, we compared Eq. (7a) to the next model, where both Cn and DI are added to age as predictors of Ci,

$$\begin{aligned}\text{Ci} = d_0 + d_1 * \text{AGE} + d_2 * \text{AGE}^2 + f_1 * \text{DI} \\ + f_2 * \text{DI}^2 + g_1 * \text{Cn} + g_2 * \text{Cn}^2\end{aligned} \tag{7b}$$

Adding Cn did not significantly increase the amount of explained variance beyond that accounted for by age and DI in Eq. (7a) [$F(2, 5) = 0.6132$, $MSE = 0.0001329$, $p > .5$] but adding DI to age and Cn did marginally improve the fit [$F(2, 5) = 5.53$, $MSE = 0.0001329$, $p = .054$].

This analysis supports the notion that DI has a direct effect on Stroop tests. First, both Cn and DI significantly reduced the error sum of squares for Ci, once the effects of age had been controlled for. Therefore, not all of the variance in Ci (and as a result in SI) across age groups can be explained by age. Second, DI differences across age groups were the better predictors of residual differences in Ci, after controlling for age. We conclude that, in this sample, DI has a significant impact on Stroop effects above and beyond age effects.

DI as a Potential Mediator of Age Effects on Ci

In this section, we investigated whether DI can serve as a mediator between age and Ci. First, we used the relationship between DI and age to predict DI from participants' age. Recall that DI was found to increase non-linearly with age (Figure 5A). We modeled this non-linear relationship with an exponential function (Eq. 3a). Recall that Ci was found to increase linearly with DI (Figure 5B). Therefore, at the next step, we used DI scores predicted from age, to predict latencies for Ci in a linear function. The indirect path from age to DI and from DI to Ci is depicted by

$$\begin{cases} \mathrm{DI} = \beta + \alpha * e^{\gamma * \mathrm{AGE}} \\ \mathrm{Ci} = d_0 + d_1 * \mathrm{DI} \rightarrow \\ \mathrm{Ci} = d_0 + d_1 * (\beta + \alpha * e^{\gamma * \mathrm{AGE}}) = (d_0 + d_1 * \beta) + (d_1 * \alpha) * e^{\gamma * \mathrm{AGE}} \end{cases} \quad (8)$$

The coefficients d_0 and d_1 were obtained from the least squares solution to the linear relation shown in Figure 5B. We then fit the coefficients β, α and γ to minimize the sum of the sum of squared errors (SSE) of the exponential relation between age and DI (the first line in Eq. 8) and the SSE of the mediated relationship between Ci and age (the third line in Eq. 8; all coefficients are presented in Appendix A). This indirect model accounted for most of the variance in Ci, 99.2%. Furthermore, this model was found to be virtually identical to an equivalent exponential model in which Ci is predicted directly from age (Eq. 3a, $r^2 = 99.2\%$), with very similar coefficients in both.

We also tested, in Eq. (9a), a mediated model using the quadratic relationship between DI and age (the dashed curve in Figure 5A) instead of the exponential relation.

$$\begin{cases} \mathrm{DI} = f_0 + f_1 * \mathrm{AGE} + f_2 * \mathrm{AGE}^2 \\ \mathrm{Ci} = d_0 + d_1 * \mathrm{DI} \rightarrow \\ \mathrm{Ci} = d_0 + d_1 * (f_0 + f_1 * \mathrm{AGE} + f_2 * \mathrm{AGE}^2) = (d_0 + d_1 * f_0) \\ \qquad + (d_1 * f_1) * \mathrm{AGE} + (d_1 * f_2) * \mathrm{AGE}^2 \end{cases} \quad (9)$$

Again, the coefficients d_0 and d_1 were obtained from the least squares solution to the linear relation shown in Figure 5B. We then fit the coefficients f_0, f_1, and f_2 to minimize the same quantity as in Eq. (8). This indirect model accounted for 98.3% of the variance. Again, the model was found to be very similar to the direct model (r^2 = 98.3%), with similar coefficients (see Appendix A). We conclude that DI may potentially serve as a mediator between age and Stroop performance. This result suggests a sensory source to age-related increases in Stroop effects.

Testing Models of Cognitive Slowing Coupled with Multiplicative Functions of Task Difficulty

In this section, we investigated whether multiplicative functions of task difficulty (as a part of age-related cognitive slowing models) could fit the data from Van der Elst et al. (2006). We first tested the multiplicative model suggested by Cerella and Halle (1994) depicted in Eq. (3b). Note that, according to this model, RTs for different tasks should be linearly related to each other. This means that with three tasks (subscript i = 1, 2, 3):

$$\begin{cases} \mathrm{RT}_1 = \mathrm{TD}_1 * \Psi(\mathrm{AGE}) \\ \mathrm{RT}_2 = \mathrm{TD}_2 * \Psi(\mathrm{AGE}) \\ \mathrm{RT}_3 = \mathrm{TD}_3 * \Psi(\mathrm{AGE}) \end{cases} \tag{10}$$

That is, reaction times on tasks 2 and 3 should be multiplicatively related to reaction times on task 1. To check whether this was indeed the case, we approximated Ψ(AGE) as a quadratic function[3] and tested the null hypothesis that,

$$\begin{cases} \mathrm{RT}_1 = \mathrm{TD}_1 * (d_0 + d_1 * \mathrm{AGE} + d_2 * \mathrm{AGE}^2) \\ \mathrm{RT}_2 = \mathrm{TD}_2 * (d_0 + d_1 * \mathrm{AGE} + d_2 * \mathrm{AGE}^2) \\ \mathrm{RT}_3 = \mathrm{TD}_3 * (d_0 + d_1 * \mathrm{AGE} + d_2 * \mathrm{AGE}^2) \end{cases} \tag{11a}$$

We arbitrarily set TD_1 to 1, and found values for TD_2, TD_3, d_0, d_1, and d_2 that minimized the sum of squared differences between the obtained and

[3]We could have expressed Ψ(AGE) as an exponential function of age. However, given the close correspondence between the quadratic version in Eq. (11a), and its equivalent exponential version (Eq. 3b), and the fact that it is easier to test the null hypothesis using the quadratic form, we chose to test the null hypothesis in quadratic form. Figure 3 shows the degree of correspondence.

predicted reaction times (see Appendix B, for a more detailed discussion of this model). This model was compared to a full model in which

$$\begin{cases} \mathrm{RT}_1 = d_{1,0} + d_{1,1} * \mathrm{AGE} + d_{1,2} * \mathrm{AGE}^2 \\ \mathrm{RT}_2 = \mathrm{TD}_2 (d_{2,0} + d_{2,1} * \mathrm{AGE} + d_{2,2} * \mathrm{AGE}^2) \\ \mathrm{RT}_3 = \mathrm{TD}_3 (d_{3,0} + d_{3,1} * \mathrm{AGE} + d_{3,2} * \mathrm{AGE}^2) \end{cases} \quad (11b)$$

We then tested the null hypothesis where $d_{1,0} = d_{2,0} = d_{3,0}$, $d_{1,1} = d_{2,1} = d_{3,1}$, and $d_{1,2} = d_{2,2} = d_{3,2}$ which was rejected [$F(6, 27) = 23.58$, MSE = 0.0079233 $p < .000001$, see Appendix B]. Hence, this multiplicative model cannot account for the data from Van der Elst et al. (2006). Figure 6 plots the predicted relationship between task difficulty and age for the multiplicative model (Eq. 11a, dashed line) for Rn, Cn, and Ci. Also shown is the predictions of the full model in which growth rates as a function of age differ across tasks (Eq. 11b, solid line). Figure 6 shows that the multiplicative model overestimates the growth rate for Rn. In fact, the multiplicative model provides a poorer fit to Rn than does the mean of Rn. Figure 6 also shows that the multiplicative model overestimates the growth rate for Cn, and underestimates the growth rate for Ci, and accounts for 60% of the variance in Cn, and 94.9% of the variance in Ci. On the other hand, the full model accounts for 97.3, 98 and 98.3% of the variances, respectively.

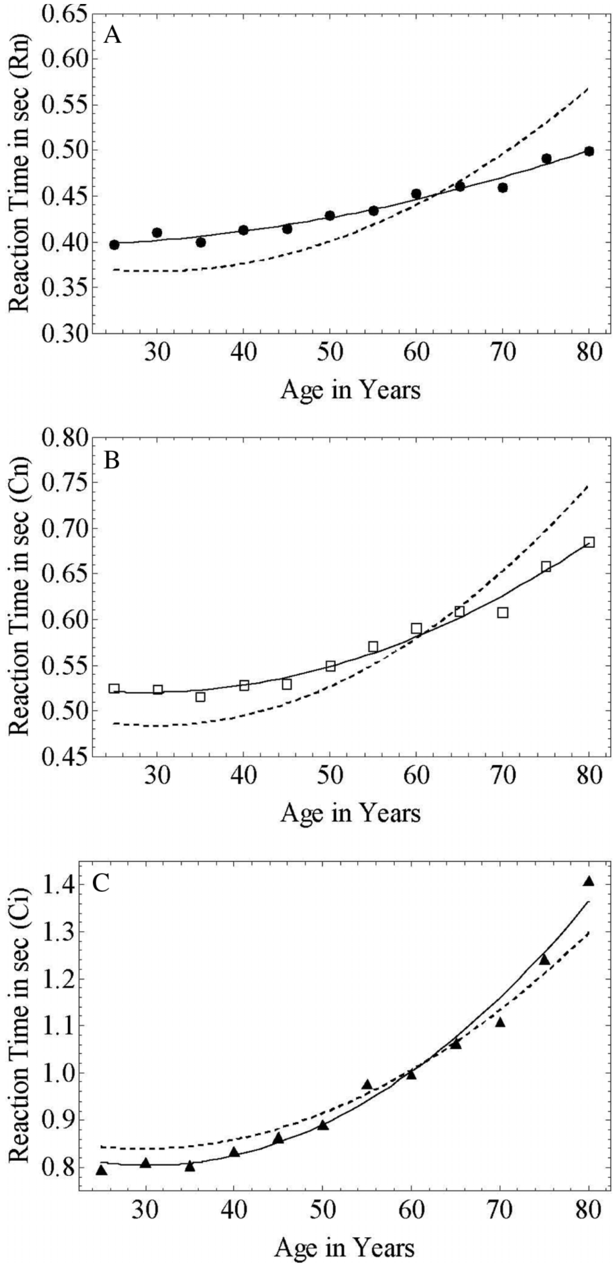
We also considered a second model of multiplicative relation of task difficulty on RTs (e.g., Verhaeghen & Cerella, 2002), namely,

$$\mathrm{RT} = \mathrm{TD} * \alpha * e^{\gamma * \mathrm{AGE}} + \beta \quad (12)$$

This model could easily fit the cross-sectional data in each of the tasks, separately. However, Eq. (12) implies that the slope of the line in the Brinley plot should be[4] $e^{\gamma * (\mathrm{Age}_{\mathrm{Old}} - \mathrm{Age}_{\mathrm{Young}})}$. The best-fitting exponential equation to Ci, according to this model, would yield a slope of 9.4 in the Brinley plot analysis (Brinley slopes for Rn and Cn would be 3.2 and 5.9, respectively), a value that is much higher than any slope values found in Brinley analyses (Verhaeghen & Cerella, 2002). Hence, this version of a multiplicative task difficulty model is also rejected by the Van der Elst et al. (2006) data.

[4]We assume the multiplicative generalized slowing model depicted in Eq. (12). Hence, the slope in the Brinley plot is $\frac{\mathrm{RT}_{\mathrm{O,TD}_2} - \mathrm{RT}_{\mathrm{O,TD}_1}}{\mathrm{RT}_{\mathrm{Y,TD}_2} - \mathrm{RT}_{\mathrm{Y,TD}_1}} = \frac{[\mathrm{TD}_2 * \alpha * e^{\gamma * \mathrm{AGE}_{\mathrm{O}}} + \beta] - [\mathrm{TD}_1 * \alpha * e^{\gamma * \mathrm{AGE}_{\mathrm{O}}} + \beta]}{[\mathrm{TD}_2 * \alpha * e^{\gamma * \mathrm{AGE}_{\mathrm{Y}}} + \beta] - [\mathrm{TD}_1 * \alpha * e^{\gamma * \mathrm{AGE}_{\mathrm{Y}}} + \beta]} = \frac{e^{\gamma * \mathrm{AGE}_{\mathrm{O}}}}{e^{\gamma * \mathrm{AGE}_{\mathrm{Y}}}}$, where subscripts O and Y refer to groups of older and younger adults, respectively.

FIGURE 6. The predictions of a model which allowed for different growth rates as a function of age (solid line, Eq. 11b) for reading neutral words (Rn, panel A), for color-naming neutral stimuli (Cn, panel B), and for color-naming incongruent words (Ci, panel C) versus a model in which the reaction times on the three tasks were assumed to be multiplicatively related (dashed line, Eq. 11a). Data from Van der Elst et al. (2006).



In sum, a cross-sectional analysis of the life-span data taken from Van der Elst et al. rejects multiplicative models of cognitive slowing as the sole explanation of age-related differences in Stroop effects.

GENERAL DISCUSSION

In an attempt to identify the factors responsible for age-related changes in Stroop performance, we analyzed the results from 13 studies which compared groups of younger and older adults on color-word Stroop tasks, and one extensive cross-sectional lifespan study with a large number of participants in each age group (Van der Elst et al., 2006). The results from the cross-lab analysis of 13 studies provided support for the hypothesis that reaction times on all tasks in a Stroop test were multiplicatively related one to the other and for a hypothesis that links dimensional imbalance (DI, the difference in latency between color-naming and reading neutral stimuli) to latencies for color-naming incongruent words (Ci). An analysis of results taken from Van der Elst et al.'s (2006) cross-sectional study further indicated that DI not only could mediate the effects of age on Stroop, but also contributed significantly to Stroop performance after controlling for the effects of age. Moreover, a model in which task difficulty interacted with cognitive slowing in a multiplicative fashion was rejected by the cross-sectional data as the complete explanation of age-related changes in Stroop performance. Below, we explore possible reasons why a cross-lab analysis of 13 studies is consistent with a multiplicative hypothesis of generalized slowing whereas an analysis of the cross-sectional data from Van der Elst et al. is not.

A Cross-Lab Analysis of 13 Studies that Compared Performance of Age Groups

In an analysis of data taken from 13 different studies of paired groups of younger and older adults, we found that the latencies for color-naming neutral words (Cn) were slowed by aging approximately twice as much as latencies for reading (Rn), resulting in larger DI scores for seniors than for younger adults in all 13 studies. These results suggest that older adults, when asked to name the font color of an incongruent color-word, might have to inhibit the lexical processing and/or response to the word for a longer time than younger adults, to permit them to name the font color. Next, we conducted a Brinley analysis, plotting latencies of older adults as a function of latencies for younger adults. The results of the Brinley analysis were compatible with a generalized slowing hypothesis coupled with a multiplicative function of task difficulty on reaction times. A single linear regression line was found to fit Cn, Ci, and Rn, implying that the same slowing mechanism may account for age differences in all three tasks. Hence, these results were consistent with the hypothesis that Stroop interference ($SI = Ci - Cn$) could be a simple by-product of generalized slowing.

An Analysis of a Single Life-Span Study Comparing 12 Age Groups Cross-sectionally

The analysis of a single life-span study (Van der Elst et al., 2006, with 1788 participants in 12 age groups), however, rejected generalized cognitive slowing as the single source for age-related changes in Stroop effects. Latencies for all three tasks, Ci, Cn and Rn, increased non-linearly with age, but with significantly different growth rates (a higher rate for Ci than for Cn, which in turn, had a higher rate than for Rn). Hence, generalized slowing, coupled with multiplicative functions relating tasks of different difficulties, could not explain reaction times on the three tasks together. As a result, generalized slowing cannot fully account for age differences in Stroop interference (the advantage of Ci over Cn) or for DI (the advantage of Rn over Cn). After controlling for the effects of age on Ci in the van der Elst et al. study, we also found DI to be the better predictor of Ci (better than Rn or Cn). Finally, support was found for a model in which the effect of age on Ci is mediated by DI. This indirect path from age to DI and from DI to Ci was as predictive as the direct path from age to Ci. This model implies that DI not only can account for some of the variance in Ci, once the effects of age have been removed, but it can also mediate age effects on Ci. Therefore, dimensional imbalance appears to play an important role in age-related changes in Stroop effects.

A Sensory Source of Age-Related Changes in Stroop Effects?

There is ample evidence in the literature that color-naming performance slows at a faster rate with age than does reading (e.g., Salthouse & Meinz, 1995; see also Cohn, Dustman, & Bradford, 1984). We propose a possible sensory source for this increase in dimensional imbalance with age, namely, sensory deterioration in color vision with age. In the introduction, we reviewed a number of studies which show that color vision deteriorates with age (e.g., Werner & Steele, 1988). On the other hand, other studies describe the relative stability of reading speed over the life-span, citing the extensive experience for reading and the optimization of font as possible reasons (e.g., Akutsu et al., 1991). Hence, it is reasonable to attribute age-related increases in DI to age-related declines in color vision. An analysis of the Van der Elst et al. (2006) data further indicates that the non-linear increase in DI scores with age cannot be attributed to generalized cognitive slowing, since the functions relating Rn and Cn to age are not multiplicatively related to each other. In sum, we suggest that age-related changes in color vision are primarily responsible for age-related changes in DI, and that it is the dimensional imbalance between access to the font's color versus access to the lexical meaning of the word that accounts, in part, for the age-related variance in the ability to name the font color of incongruent color-words. That is, a sensory factor, DI, contributes to age-related changes in Stroop effects.

It is worth noting that the results from our analyses of color-word Stroop studies are consistent with the information degradation explanation (Lindenberger & Baltes, 1994; Schneider & Pichora-Fuller, 2000) of how sensory declines might affect cognitive performance. According to this hypothesis, declines in cognitive performance could arise because the information delivered by the sensory systems becomes degraded with age. In the case of the Stroop test, age-related losses in color vision could be delaying access to the font color of the word, thereby increasing the load on working memory (Melara & Algom, 2003), because the lexical response to the word will have to be inhibited for a longer period of time. Hence, the greater susceptibility of older adults to Stroop interference could be due, at least in part, to the fact that sensory declines directly lead to a longer period of inhibition. Note, of course, that it does not entirely eliminate the inhibitory deficit explanation of Stroop interference (Hasher & Zacks, 1988). It could be that, in addition to the longer inhibition period, older adults are also less able to inhibit irrelevant material than are younger adults. Further experiments will be required to separate the relative contributions of sensory decline versus age-related deficits in inhibition.

Comparing Cross-Lab Analyses with Cross-Sectional Analyses

Our cross-lab analysis of the results from 13 studies of Stroop effects on younger and older adults was found to be consistent with a generalized slowing hypothesis coupled with a multiplicative explanation of the differences in task difficulty. Specifically, a single straight line in a Brinley plot was able to account for all of the data in each of the three tasks. These results are inconsistent with our cross-sectional analysis of the Van der Elst et al. (2006) data, in which different non-linear functions related reaction time to age for each of the three tasks. One plausible reason for this discrepancy is that there is less statistical power when studies are collapsed across labs in a Brinley plot, than when cross-sectional data are collected in a single lab using identical tasks and testing procedures for all age groups. In this section, we explore whether differences in statistical power can explain this discrepancy.

Recall that in order for all data points in a Brinley plot to fall on the same straight line in a generalized slowing account, the same function must underlie performance in each of the three tasks. Let’s examine Cerella and Halle’s (1994) model of generalized slowing coupled with a multiplicative function of task difficulty (Eq. 3b). If we now assume that there are variations in TD_i between labs due to a variety of factors, we can write,

$$\mathrm{RT}_{i,k} = \mathrm{TD}_i * \mathrm{LD}_{i,k}[\alpha * e^{\gamma \mathrm{AGE}} + \beta] \tag{13}$$

where $LD_{i,k}$ specifies the degree to which TD_i varies between labs, with subscript k representing different labs. The difference in statistical power between the two analyses ensues from Eq. (13). In the life-span study of Van der Elst et al. (2006), with 3 tasks and 12 age groups (producing 36 data points), there are only three task difficulty parameters (one for each task). However, when only two age groups are tested, with 3 tasks across 13 different labs (producing 39 data points in a Brinley plot), there are 39 different task difficulty parameters (3 for each of the 13 labs). Clearly, the power for detecting a specific research hypothesis (different α_i, β_i, and γ_i for each task) has to be less for the Brinley plot than for a cross-sectional study having a similar number of data points. Moreover, note that there were more participants per age group in the Van der Elst et al. (2006) data (an average of 149, see Table 2) than in the cross-lab data (an average of 70, see Table 1). Hence, for the same effect size in both analyses there is greater power in the cross-sectional analysis performed here. Finally, it should be noted that although the data of Van der Elst et al. (2006) clearly reject a multiplicative model of generalized slowing, it is always possible that their results are anomalous for some unknown reason or reasons. Given the large N and high levels of significance found in this study, we believe that this is a remote possibility. Nevertheless, it would be helpful to have their cross-sectional approach replicated in other languages and in other laboratories to assess the generality of their results.

Summary

We have shown that the traditional comparative analysis of two age groups (younger and older adults) across labs in a Brinley analysis, used in previous meta-analyses on Stroop effects (e.g., Verhaeghen & De Meersman, 1998 and Verhaeghen & Cerella, 2002), may not be sensitive and powerful enough to detect the full range of age-related changes evident in a life-span analysis. Specifically, a cross-sectional approach was consistent with the hypothesis that age-related cognitive changes on the Stroop task were mediated by sensory declines in color vision. The failure of Brinley plots to reveal the consequences of this sensory decline was attributed to the lack of statistical power available in comparisons of only two age groups across laboratories. We conclude that studies of age-related changes in Stroop effects must take into account sensory factors, specifically the deterioration of color vision with age. Future empirical research is needed to establish the extent to which these sensory factors, rather than cognitive factors (such as speed of processing, or age-related inhibitory deficits), are responsible for age-related changes in Stroop interference.

ACKNOWLEDGEMENTS

This study was supported by a strategic training grant – Communication and Social Interaction in Healthy Aging, and a group grant on Sensory and Cognitive Aging, both funded by the Canadian Institutes of Health Research, and a grant from the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank Huiwen Goy, Linh Nguyen, and Wu Yan (Lulu) Li for their assistance in retrieving the relevant studies.

Original manuscript received May 26, 2008
Revised manuscript accepted February 25, 2009
First published online May 28, 2009

REFERENCES

Akutsu, H., Legge, G. E., Ross, J. A., & Schuebel, K. J. (1991). Psychophysics of reading – X. Effects of age-related changes in vision. *Journal of Gerontology, 46*, 325–331.

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24*, 557–572.

Anstey, K. M., Dain, S., Andrews, S., & Drobny, J. (2002). Visual abilities in older adults explain age-differences in Stroop and fluid intelligence but not face recognition: Implications for the vision-cognition connection. *Aging, Neuropsychology, and Cognition, 9*(4), 253–265.

Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition, 14*(2), 155–167.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*(4), 859–873.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In: A. T. Welford, & J. E. Birren (Eds), *Behavior, aging and the nervous system* (pp. 114–149). Springfield, IL: C. C. Thomas.

Cerella, J. (1990). Aging and information processing rate. In J. E. Birren & K. W. Schaie (Eds), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.

Cerella, J., & Hale, S. (1994). The rise and fall of information processing rates over the life span. *Acta Psychologica, 86*, 109–197.

Cohn, N. B., Dustman, R. E., & Bradford, D. C. (1984). Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology, 40*, 1244–1250.

Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test on childhood, adulthood, and aging. *Journal of Genetic Psychology, 100*, 47–53.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 470–484.

Farnsworth, D. (1943). The Farnsworth–Munsell 100-hue and dichromatous tests for colour vision. *Journal of the Optical Society of America, 33*, 568–578.

Golden, C. J. (1978). *The Stroop Color and Word Test*. Chicago: IL: Stoelting Company.

Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously relevant information. *Psychology and Aging, 6*, 587–594.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Houx, P. J., Jolles, J., & Vreeling, F. W. (1993). Stroop interference: Aging effects assessed with Stroop color-word test. *Experimental Aging Research, 19*, 209–224.

Kieley, J. M., & Hartley, A. A. (1997). Age-related equivalence of identity suppression in the Stroop color-word task. *Psychology and Aging, 12*(1), 22–29.
Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576–588.
Klein, M., Ponds, R. W. H. M., Houx, P. J., & Jolles, J. (1997). Effect of test duration on age-related differences in Stroop interference. *Journal of Clinical & Experimental Neuropsychology, 19*, 77–82.
Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 9*, 339–355.
Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 978–992.
MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.
McDowd, J. M., & Shaw, R. J. (2000). Attention and aging: A functional perspective. In F. I. M. Craik, & T.A. Salthouse (Eds), *The handbook of aging and cognition* (2nd ed., pp. 221–292). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Melara, R. D. & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*, 422–471.
Monsell, S., Taylor T. J., & Murphy K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition, 29*(1), 137–151.
Nguyen-Tri, D., Overbury, O., & Faubert, J. (2003). The role of lentricular senescence in age-related color vision changes. *Investigative Ophthalmology & Visual Science, 44*(8), 3698–3704.
Rodriguez-Aranda, C. (2003). Reduced writing and reading speed and age-related changes in verbal fluency tasks. *The Clinical Neuropsychologist, 17*(2), 203–215.
Rodriguez-Aranda, C., & Sundet, K. (2006). The frontal hypothesis of cognitive aging: Facture structure and age effects on four frontal tests among healthy individuals. *The Journal of Genetic Psychology, 167*(3), 269–287.
Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failure of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 515–537.
Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journals of Gerontology: Psychological Sciences, 50*(6), 297–306.
Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik, & T.A. Salthouse (Eds), *The handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 461–479.
Stroop, J. R. (1935). Studies of interference in serial-verbal reaction. *Journal of Experimental Psychology, 18*, 643–662.
Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *The Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.
Troyer, A. K., Leach, L., & Strauss, E. (2006). Aging and response inhibition: Normative data for the Victoria Stroop test. *Aging, Neuropsychology, and Cognition, 13*(1), 20–35.
Uttl, B., & Graf, P. (1997). Color-word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology, 19*(3), 405–420.

Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2006). The Stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*(1), 62–79.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging*, *13*(1), 120–126.

Verhaeghen, P., & Cerella, J. (2002).Aging, executive control, and attention: a review of meta-analyses. *Neuroscience and Biobehavioral Reviews*, *26*, 849–857.

Werner, J. S., & Steele, V. G. (1988). Sensitivity of human foveal color mechanisms throughout the life span. *Journal of the Optical Society of America. A, Optics and image science*, *5*(12), 2122–2130.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging*, *13*(2), 206–217.

APPENDIX A

The coefficients of the best fitting exponential model, in which DI mediates between Ci and AGE (Eq. 8), were: β = 0.108159, α = 0.00143279, γ = 0.0499839, d_0 = –0.13471, d_1 = 8.20778; the coefficients of the exponential model, which provided the best direct fit of age and Ci, $\mathrm{Ci} = A + B * e^{\gamma * \mathrm{AGE}}$, were: A = 0.752285, B = 0.0119393, γ = 0.04980.

The coefficients of the best fitting quadratic model, in which DI mediates between CI and AGE (Eq. 9), were: f_0 = 0.140685, f_1 = –0.00171755, f_2 = 0.00002803337, d_0 = –0.13471, d_1 = 8.20778; the coefficients of the quadratic model, which provided the best direct fit of AGE and Ci, $\mathrm{Ci} = F_0 + F_1 * \mathrm{AGE} + F_2 * \mathrm{AGE}^2$, were: F_0 = 1.017, F_1 = –0.0139851, F_2 = 0.00022915.

APPENDIX B

We assume the multiplicative generalized slowing model depicted in Eq. (3b). In a Brinley plot, we plot $\mathrm{RT_O}$, average reaction times for the groups of older adults as a function of $\mathrm{RT_Y}$, reaction times of groups of younger adults, specifically,

$$\begin{cases} \mathrm{RT_O} = \mathrm{TD} * \Psi(\mathrm{AGE_O}) = \mathrm{TD} * (\alpha * e^{\gamma * \mathrm{AGE_O}} + \beta) \\ \mathrm{RT_Y} = \mathrm{TD} * \Psi(\mathrm{AGE_Y}) = \mathrm{TD} * (\alpha * e^{\gamma * \mathrm{AGE_Y}} + \beta) \end{cases} \quad (14)$$

Note, that a fixed TD defines one point in a Brinley plot, and that variations in TD have the same multiplicative effect on $\mathrm{RT_O}$ and $\mathrm{RT_Y}$, constraining the Brinley function to be linear with an intercept of zero. By having two values of task difficulty, we define two points on a Brinley plot: ($RT_{Y,TD1}$, $RT_{O,TD1}$), and ($RT_{Y,TD2}$, $RT_{O,TD2}$). Hence, the slope in the Brinley plot is

$$\begin{cases} \dfrac{RT_{O,TD_2} - RT_{O,TD_1}}{RT_{Y,TD_2} - RT_{Y,TD_1}} \\ \quad = \dfrac{TD_2 * [\alpha * e^{\gamma * AGE_O} + \beta] - TD_1 * [\alpha * e^{\gamma * AGE_O} + \beta]}{TD_2 * [\alpha * e^{\gamma * AGE_Y} + \beta] - TD_1 * [\alpha * e^{\gamma * AGE_Y} + \beta]} \\ \quad = \dfrac{\alpha * e^{\gamma * AGE_O} + \beta}{\alpha * e^{\gamma * AGE_Y} + \beta} \\ \text{Slope} = \dfrac{RT_O}{RT_Y} \end{cases} \tag{15}$$

Notice that the slope of the Brinley plot is solely determined by α, β, and γ, and the ages of the two groups. Therefore, if two different tasks follow a multiplicative age-related slowing model, all points in a Brinley plot will fall on the same linear line.

When we expressed how RTs change with age across the life-span, as a quadratic function, we were able to fit values to the null and research hypotheses specified by Eqs (11a) and (11b) respectively. If we set $TD_1 = 1$, the full model (Eq. 11b) has 9 parameters ($d_{1,0}$, $TD_2 * d_{2,0}$, $TD_3 * d_{3,0}$, $d_{1,1}$, $TD_2 * d_{2,1}$, $TD_3 * d_{3,1}$, $d_{1,2}$, $TD_2 * d_{2,2}$, and $TD_3 * d_{3,2}$), the reduced model (Eq. 11a) has 5 parameters (TD_2, TD_3, d_0, d_1, d_2). To test whether increasing the number of parameters from 5 to 9 produced a significant reduction in the sum of squared errors between the predicted and obtained values, we first fixed the values of TD_2 and TD_3 as $Mean[RT_2]/Mean[RT_1]$ and $Mean[RT_3]/Mean[RT_1]$, respectively. With TD_2 and TD_3 fixed at these values, we found the values of d_0, d_1, and d_2 that minimized the sum of squared differences between the obtained reaction times (from Van der Elst et al., 2006) and those predicted by the best-fitting reduced model, as well as the values of $d_{1,0}$, $d_{2,0}$, $d_{3,0}$, $d_{1,1}$, $d_{2,1}$, $d_{3,1}$, $d_{1,2}$, $d_{2,2}$, and $d_{3,2}$ that minimized the sum of squared differences between the obtained reaction times and those predicted by the best-fitting full model. The error sum of squares for the reduced model was then noted for these fixed values of TD_2 and TD_3. We then varied the values of TD_2 and TD_3 around the starting values to obtain the smallest error sum of squares for the reduced model, and those were the values we used for the *F*-test reported here. Under those conditions, the full model has 9 free parameters, and the reduced model has 3 free parameters.

Hence the degrees of freedom in the *F*-test were set to 9–3 = 6 in the numerator, and 36–9 = 27 in the denominator. The best fitting values of TD_2 and TD_3 that minimized the sum of squared errors in the reduced model were 1.32 and 2.28, respectively. The best fitting values for the reduced model were 0.43736, –0.004694 and 0.0000792, for d_0, d_1, and d_2, respectively. The best fitting values for the full model for these values of TD_2 and TD_3 are: 0.40143, 0.43377, and 0.44545 for $d_{1,0}$, $d_{2,0}$, $d_{3,0}$, respectively; –0.000725, –0.002678, –0.006125 for $d_{1,1}$, $d_{2,1}$, $d_{3,1}$, respectively; and 0.0000245, 0.0000469 and 0.0001004 for $d_{2,1}$, $d_{2,2}$, $d_{2,3}$, respectively. Note that one could argue that the appropriate degrees of freedom for the reduction in sum of squares in going from the reduced to the full model is 9–5 = 4, because we actually fit 5 parameters in the reduced model. However, two of the coefficients enter into the reduced model in a multiplicative rather than in an additive fashion. Hence, it is not clear what the appropriate number of degrees of freedom are for the reduced model. Using 9–3 = 6 is a more conservative criterion for significance because, for the same Type I error rate, the critical value of $F[6, 27] > F[4, 27]$.